

Architecture and Algorithm for a Stochastic Soft-output MIMO Detector

Kiarash Amiri, Predrag Radosavljevic, Joseph R. Cavallaro

Department of Electrical and Computer Engineering, Rice University
6100 Main St., Houston, TX 77005
{kiaa, rpredrag, cavallar}@rice.edu

Abstract—In this paper, we propose a novel architecture for a soft-output stochastic detector in multiple-input, multiple-output (MIMO) systems. The stochastic properties of this detector are studied and derived in this work, and several complexity reduction techniques are proposed to significantly reduce its cost from an architecture-implementation perspective. We also propose an efficient architecture to implement this detector. Finally, this detector is incorporated into an iterative detection-decoding structure, and through simulations, it is shown that the overall frame error rate (FER) performance and complexity is of the same order as that of the conventional K-best sphere detector.

I. INTRODUCTION

Multiple-input Multiple-output (MIMO) systems [1] can achieve both higher reliability and higher data rates compared to single-antenna wireless systems. Therefore, they have been extensively used in many different wireless systems, and have been proposed for current and upcoming wireless standards, such as cellular networks, IEEE 802.11n, WiMAX and 3GPP LTE. In all these standards, broadcasting independent streams on the antennas, commonly referred to as spatial multiplexing, plays a significant role in pushing the data rates to support higher demanding applications, such as video streaming.

Implementing these types of MIMO systems in base stations and hand-held mobile devices poses a wide range of issues, among which detection is one of the major challenges. There has been significant work on architecture-friendly techniques to reduce the complexity of near-optimum solutions for MIMO detection [2], [3], [4]. These techniques are all based on sphere detection, [5], which mimics the maximum-likelihood (ML) performance. However, there are still inevitable costly operations associated with those techniques, which consume a significant portion of the area/power of the wireless system. In order to avoid some of these computations, a novel approach has been proposed in [6], which is based on Markov Chain Monte Carlo (MCMC) technique, and takes a stochastic approach to reduce the soft detection complexity.

In this paper, we propose a novel stochastic detector which uses the general Markov chain approach. We show that the updated distribution is a Gaussian distribution with a running mean and tractable variance. We further propose various complexity reduction techniques to efficiently compute the parameters of the distribution, and propose an architecture to implement this technique. Our simulation results suggest

that the performance and computational complexity of this approach is comparable with that of common sphere detection.

This paper is organized as follows: Section II introduces the system model. Section III introduces the stochastic detector based on the Markov Chain Monte Carlo (MCMC) technique. An efficient architecture, suitable for ASIC implementation, is proposed in section IV. Finally, the simulation results are presented in section V.

II. SYSTEM MODEL

Let the transmitted vector, \mathbf{s} , be an $M_T \times 1$ vector with its complex elements, s_i , chosen from a set of modulation constellation with 2^{M_c} complex signal points. Then, the MIMO system model with M_T transmit antennas and M_R receive antennas can be described by

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n} \quad (1)$$

where \mathbf{H} is the $M_R \times M_T$ complex-valued channel matrix, \mathbf{n} is the $M_R \times 1$ complex noise vector, and \mathbf{y} is the $M_R \times 1$ received vector. Here, each $s_i, i = 1, \dots, M_T$ corresponds to M_c -length bit sequence of x_i s where M_c is the number of bits per modulation symbol. Throughout the paper, we assume that the receiver has perfect CSI; i.e. the channel matrix, \mathbf{H} , is fully known in the receiver. We are only concerned with full-rate spatial multiplexing, with $M_T \leq M_R$, which uses all transmit antennas for transmission.

Figure 1 shows the overall system architecture. The transmitter consists of an outer error correction encoder, e.g. Turbo or LDPC encoder, of rate R ; followed by a constellation mapper which maps the encoded bits to complex modulation symbols, e.g. 16-QAM or 64-QAM. The receiver uses an iterative detection/decoding scheme [7] which is based on exchanging soft information between detector and soft-in/soft-out decoder in a feedback loop until reliable decoded bits are computed. The soft information, typically Log-likelihood Ratio (LLR), passed from the detection block to the decoding block is obtained by

$$L_D(x_k|\mathbf{y}) = \ln \frac{P[x_k = +1|\mathbf{y}]}{P[x_k = -1|\mathbf{y}]} \quad (2)$$

where $k = 0, \dots, M_T \cdot M_c - 1$. This soft information is updated in the decoder and fed back into detector. Multiple cycles of exchanging soft information between the detector and decoder

would eventually lead to more reliable soft information, which will be used by the decoder, in the last iteration, to hard-decode more reliably.

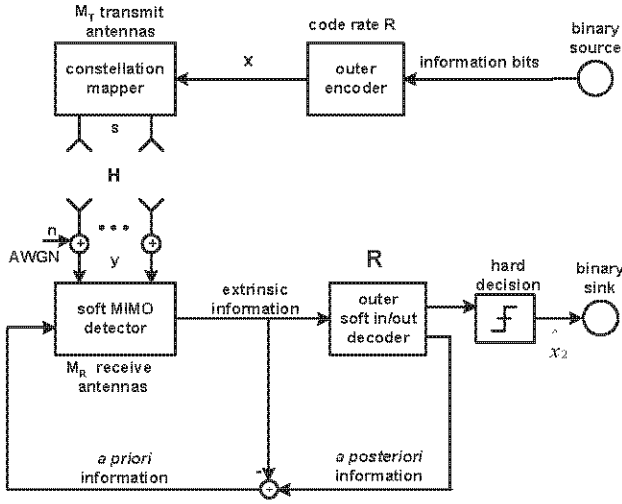


Fig. 1. MIMO transceiver, iterative detection and decoding at the receiver.

Soft information can be generated using a list of possible vector candidates. Once this list is generated, LLR values of Eq. (2) are computed and passed to the decoder [7]:

$$L_E(x_k|y) \approx \frac{1}{2} \max_{\mathbf{x} \in \mathcal{L} \cap \mathbf{X}_{k+1}} \left\{ -\frac{1}{\sigma^2} \|\mathbf{y} - \mathbf{H} \cdot \mathbf{s}\|^2 + \mathbf{x}_{[k]}^T \cdot \mathbf{L}_{A,[k]} \right\} - \frac{1}{2} \max_{\mathbf{x} \in \mathcal{L} \cap \mathbf{X}_{k-1}} \left\{ -\frac{1}{\sigma^2} \|\mathbf{y} - \mathbf{H} \cdot \mathbf{s}\|^2 + \mathbf{x}_{[k]}^T \cdot \mathbf{L}_{A,[k]} \right\}, \quad (3)$$

where \mathcal{L} is the list of possible vectors, $\mathbf{x}_{[k]}$ is the sub-vector of \mathbf{x} obtained by omitting the k -th bit x_k , $\mathbf{L}_{A,[k]}$ is the vector of all *a priori* probabilities L_A for transmitted vector \mathbf{x} obtained by omitting $L_A(x_k)$, σ^2 is the noise variance, \mathbf{X}_{k+1} is the set of $2^{M \cdot M_C - 1}$ bits of vector \mathbf{x} with $x_k = +1$, while \mathbf{X}_{k-1} is similarly defined.

A. Soft Output Sphere Detector

One straightforward way to generate the list, \mathcal{L} in Eq. (3), is using sphere detection as proposed in [7]. Note that when using a sphere detector; rather than picking the best candidate, a list of *desirable* candidates, i.e. generally the ones with smaller distances, are chosen and added to the list \mathcal{L} . Different architectures and implementations have been proposed for such soft sphere detectors in [3], [4] and [8]. They all require significant pre-processing effort, which usually includes various channel factorizations, such as QR decomposition or Cholesky factorization.

It is also worth noting that sphere detection was originally designed to alleviate the complexity of integer least-squares problems [5]. Since maximum-likelihood (ML) solution in the presence of a wireless channel is an integer least-squares problem, sphere detection is a suitable choice for wireless

receivers which try to detect the *uncoded* data, and achieve a performance close to optimum, i.e. ML, performance. However, for a vast majority of current wireless transceivers, the error correction encoder/decoder pair is a crucial integral block, see Figure 1; and for such systems, a hard decision symbol (something sphere detection is built for) is not the target; rather, soft reliability information should be computed to aid the decoder in decoding the data successfully.

Therefore, rather than using a technique that was originally designed for hard detection in uncoded systems and extending it to coded systems, it is worthwhile to consider a new approach to computing the soft reliability information.

B. Markov Chain Monte Carlo (MCMC) Technique for List Generation

In order to generate the list and compute the soft information of Eq. (3), while avoiding the computationally expensive sphere detector, a stochastic approach has been proposed in [6]. In this technique, each transmit antenna symbol is drawn out of a distribution function.

The algorithm steps for this technique are summarized as follows:

Initialize a random $\mathbf{s}^{(-N_b)}$;

For $n = -N_b + 1$ to $n = N_s$:

- Draw sample $s_1^{(n)}$ from $P(s_1 | s_2^{(n-1)}, \dots, s_{M_T}^{(n-1)}, \mathbf{y})$,
 - Draw sample $s_2^{(n)}$ from $P(s_2 | s_1^{(n-1)}, s_2^{(n-1)}, \dots, s_{M_T}^{(n-1)}, \mathbf{y})$,
 - ...
 - Draw sample $s_{M_T}^{(n)}$ from $P(s_{M_T} | s_1^{(n)}, \dots, s_{M_T-1}^{(n)}, \mathbf{y})$.
-

The preceding algorithm can be proceeded in parallel independent paths, each called one sampler unit; thus, for an average fixed list size, reducing the delay required to finish the overall procedure. Furthermore, each sampler would generate a list of highly correlated candidates; therefore, using multiple parallel samplers would increase the probability of choosing from less correlated sets of candidates, which, in turn, would increase the reliability of the soft information passed to the decoder. Throughout the paper, we denote the number of parallel units by N .

III. STOCHASTIC SOFT-OUTPUT MIMO DETECTOR

We show here that assuming additive white Gaussian noise, all the above probability density functions can be written as follows,

$$\begin{aligned} \text{pdf}_j^{(n)} &= P(s_j | s_1^{(n)}, s_2^{(n)}, \dots, s_{j-1}^{(n)}, s_{j+1}^{(n-1)}, \dots, s_{M_T}^{(n-1)}, \mathbf{y}, \mathbf{H}) \\ &= K \exp \left\{ -\frac{\|\mathbf{y} - \mathbf{H}\mathbf{s}\|^2}{2\sigma^2} \right\} \\ &= K \exp \left\{ -\frac{\sum_i |y_i - \mathbf{h}_i^T \mathbf{s}|^2}{2\sigma^2} \right\} \end{aligned}$$

$$\begin{aligned}
&= K \exp \left\{ -\frac{\sum_i |y_i - h_{ik}s_k - \sum_{j \neq k} h_{ij}s_j|^2}{2\sigma^2} \right\} \\
&= K \exp \left\{ -\frac{\sum_i |y_i - h_{ik}s_k - \sum_{j \neq k} h_{ij}s_j|^2}{2\sigma^2} \right\} \\
&= K_2 \exp \left\{ -\frac{\sum_i |h_{ik}s_k|^2 - 2\Re \left\{ s_k \sum_i h_{ik}(y_i^* - \sum_{j \neq k} h_{ij}^* s_j^*) \right\}}{2\sigma^2} \right\} \\
&= K_3 \exp \left\{ -\frac{\left| s_k - \frac{\sum_i h_{ik}(y_i^* - \sum_{j \neq k} h_{ij}^* s_j^*)}{\sum_i |h_{ik}|^2} \right|^2}{2\sigma^2 / \sum_i |h_{ik}|^2} \right\}. \quad (4)
\end{aligned}$$

where \mathbf{h}_i is the i -th row of the channel matrix, \mathbf{H} ; superscripts $*$ and T indicate scalar conjugate and vector transpose conjugate respectively, and K_i s are proper coefficients to normalize the distribution at each step.

Thus, the distribution from which the sample is drawn at each step is a Gaussian random variable with mean, α , and variance, β^2 , given as

$$\alpha = \frac{\sum_i h_{ik}^*(y_i - \sum_{j \neq k} h_{ij}s_j)}{\sum_i |h_{ik}|^2}, \quad (5)$$

$$\beta^2 = \frac{\sigma^2}{\sum_i |h_{ik}|^2}. \quad (6)$$

Assuming that a Gaussian random variable with zero mean and variance of one is given in advance, the new random variable can be obtained from

$$s_j^{(n)} = \{\Re\{\alpha\} + \beta \times \varphi\} + j\{\Im\{\alpha\} + \beta \times \psi\}, \quad (7)$$

where $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote the real and imaginary part operators; β and α are given in Eq. (5) and (6); and φ and ψ are two independent samples drawn from the given zero-mean uni-variance Gaussian distribution.

IV. ARCHITECTURAL CONSIDERATIONS

Note that the mean, Eq. (5), changes in each step. Therefore, computation of the mean can be quite expensive consisting of various additions, multiplications and divisions. In this section, we propose architecture oriented techniques to reduce this complexity.

Revisiting (5) and (6) suggests the following modifications:

- **Fixed Multiplications:** Even though the drawn sample can be any real numbers, they can be mapped to one of the modulation constellation points so that all the $h_{ij}s_j$ multiplications change to shift-add operations.
- **Avoiding Square roots/divisions:** With the current definitions of α and β in (5) and (6), division operations are required at each step to find α ; moreover, to find β , M_T number of square root calculations are required in the pre-processing stage. To avoid all these calculations, everything, including the modulation constellation, should be scaled with a $\sum_i |h_{ik}|^2$ factor. Therefore, the constellation points are now defined as

$$m_k = s_k \sum_{i=1}^{M_R} |h_{ik}|^2 \quad (8)$$

for each of the $k = 1, \dots, M_T$ transmit antennas. Therefore, the new mean and variance to be substituted in (7) are defined as:

$$\begin{aligned}
\alpha' &= \sum_i h_{ik}^*(y_i - \sum_{j \neq k} h_{ij}s_j), \\
\beta'^2 &= \sigma^2 \sum_{i=1}^{M_R} |h_{ik}|^2. \quad (9)
\end{aligned}$$

Note that this scaling depends on the instantaneous channel realization; however, a significant number of division/sqrt operations are replaced by less costly ones, i.e. multiplications. Also, notice that variance has to be adjusted only once in (9), and then being used as long as the channel estimates remain the same.

- **Iterative Computation of the Mean:** To further reduce the complexity and avoid repeating multiplications, α' can be re-written as:

$$\begin{aligned}
\alpha'^{(n+1)} &= \alpha'^{(n)} + \sum_{i=1}^{M_R} h_{ik}^* \sum_{j \neq k}^{M_T} h_{ij}s_j^{(n)} - \sum_{i=1}^{M_R} h_{ik}^* \sum_{j \neq k}^{M_T} h_{ij}s_j^{(n+1)} \\
&= \alpha'^{(n)} + \sum_{i=1}^{M_R} h_{ik}^* \left\{ \sum_{j \neq k}^{M_T} (h_{ij}s_j^{(n)} - h_{ij}s_j^{(n+1)}) \right\} \\
&= \alpha'^{(n)} + \sum_{i=1}^{M_R} \sum_{j \neq k}^{M_T} h_{ik}^* h_{ij} (s_j^{(n)} - s_j^{(n+1)}). \quad (10)
\end{aligned}$$

Using (10), the $h_{ik}^* h_{ij}$ can be computed in advance and stored; and can be re-used throughout the sample drawing steps. The $\times(s_j^{(n)} - s_j^{(n+1)})$ term in (10) is basically a sequence of shift-add operations as all the s_j s are constellation points.

Table I summarizes the resource savings through adopting the aforementioned techniques. The number of parallel samplers is denoted by N , and the number of iterations is $I = M_T(N_b + N_s)$. In order to compute the operation count, based on typical FPGA area costs, comparators are assumed to have unit complexity; adders have twice complexity as that of comparators; multipliers ten times that of addition; and finally division and square root to have 4.5 and 3.8 times that of multipliers, respectively.

TABLE I
OPERATION COUNT COMPARISON BETWEEN THE TWO APPROACHES TO COMPUTE THE MEAN AND VARIANCE IN EQ. (5)/(6) AND EQ. (9)/(10). THE TOTAL NUMBER IS FOR A 16-QAM, 4×4 SYSTEM.

	Initial stochastic detector based on (5) and (6)	Reduced stochastic detector based on (9) and (10)
comparison	$2\sqrt{w}NI$	$2\sqrt{w}NI$
addition	$M_T(M_R - 1) + N.I\{8M_R(M_T - 1) + 4M_R + 4\}$	$(2M_R - 1)M_T + 2M_R M_T(\frac{\sqrt{w}}{2} - 1) + N.I\{2M_R(M_T - 2)\} + N.M_T\{2M_R(M_T - 1) + 2(M_R - 1)\}$
square	$2M_R M_T$	$2M_R M_T$
mult	$N.I\{4M_R + 2\}$	$4(\frac{1}{2}M_R M_T(M_T - 1)) + 4M_R M_T N + 2N.I$
Division	$M_T + 2N.I$	0
Square root	M_T	0
total number of operations	18K computation operations	1.7K computation operations

Once α is computed, drawing the random variables is straightforward using (7). Figure 2 shows how this can be accomplished with memory blocks containing the instances of a zero-mean uni-variance Gaussian distribution with proper number of occurrence of different samples. The address to this memory block, which needs to be statistically uniform, is generated with a conventional linear feedback shift register (LFSR).

The overall architecture is shown in the same figure. Note that there are N parallel sampler units, and since they operate independently of each other, there is no data dependency between the samplers; thus, they can run in parallel. Each of these N sampler units generates a list \mathcal{L}_i . The final list, \mathcal{L} , required to compute Eq. (3), is generated by concatenating all these \mathcal{L}_i lists.

Finally, it is worthwhile noting that this architecture utilizes all the channel instantaneous information, i.e. $h_{ij}s$ in Eq. (9) and (10), to generate the list. However, compared to sphere detection, it does not require any explicit costly pre-processing, such as QR decomposition or Cholesky factorization to produce a similar list. A more explicit comparison on the complexity and computation cost is given in section V.

V. SIMULATION RESULTS

For the simulation phase, a 4×4 MIMO wireless system is assumed. The stochastic detector is employed in an iterative detection-decoding structure, Figure 1, where four outer iterations are performed in the detector/decoder loop. During these iterations, the updated soft information is handed between the detector and the decoder. The decoder is a $R = 0.5$ rate 1944 length LDPC decoder, and is using at most 15 inner iterations

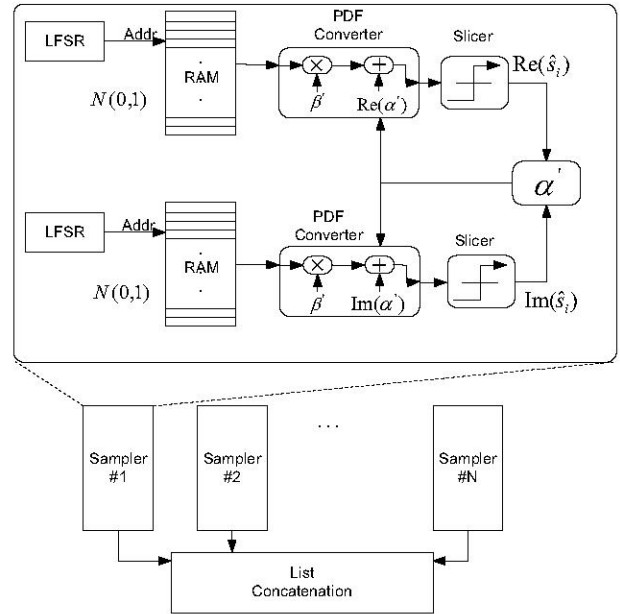


Fig. 2. MCMC architecture assuming N parallel sampler units.

between the variable and check nodes of the decoder. The detector is simulated with all the simplifications described in the previous section. The channel matrices are all independent Rayleigh fading. For the stochastic detector, $N = 8$ parallel samplers are assumed, $N_b = 5$ and $N_s = 10$, and each \mathcal{L}_i has

at most 20 vectors.

The frame error rate (FER) results comparing the overall performance of the stochastic detector with the K-best sphere detectors for both 16-QAM and 64-QAM are presented in Figure 3.

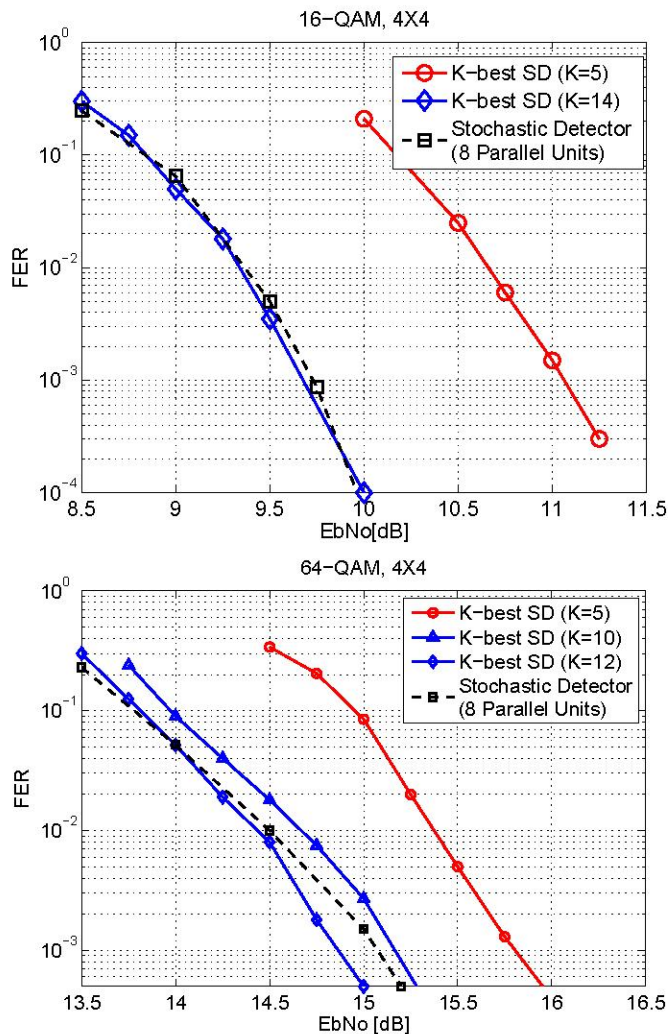


Fig. 3. FER results comparing K-best Sphere Detector (SD) with MCMC for both 16-QAM and 64-QAM systems, assuming four transmit antennas. LDPC decoder with 15 inner iterations have been used.

The number of operations for different modulation orders are shown in Figure 4. Note that for each modulation order, the complexity of K-best technique increases with higher K values of the K-best sphere detection. The complexity of higher order modulations is more significant in sphere detection due to the sorting step required in that strategy, which suggests that for higher dimension problems, the stochastic approach would show significant improvement over conventional deterministic approaches.

VI. CONCLUSION

In this work, we presented a novel architecture for an MCMC-based stochastic detector which can be used in iterative detector/decoder structures to improve the performance of wireless MIMO systems, specifically in higher dimension problems. This architecture enjoys various complexity reduction techniques, and is suitable for customized ASIC solutions. The simulation results show that the frame error rate behavior of this technique is comparable with that of sphere detection.

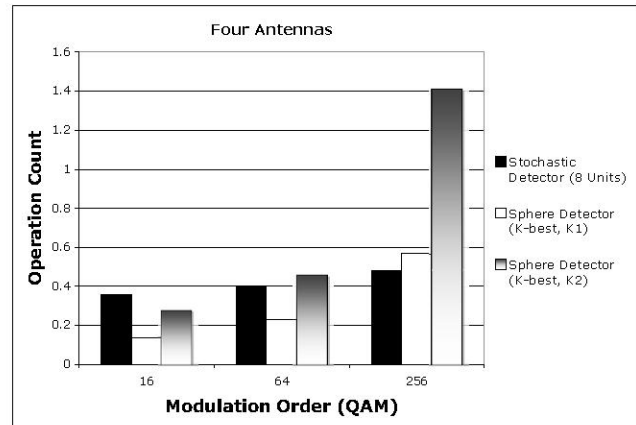


Fig. 4. Number of operations for different modulation orders, normalized by a factor of 10^5 . For the 16-QAM case, K1 and K2 correspond to 5 and 14, respectively. For 64-QAM and 256-QAM, K1 and K2 correspond to 5 and 10, respectively. Four transmit antennas are assumed.

VII. ACKNOWLEDGEMENT

This work was supported in part by Nokia Corporation, Xilinx Inc., and by NSF under grants EIA-0321266, CCF-0541363, CNS-0551692, and CNS-0619767.

REFERENCES

- [1] G. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs. Tech. Journal*, vol. 2, pp. 41–59, 1996.
- [2] A. Burg, M. Borgmann, M. Wenk, M. Zellweger, W. Fichtner and H. Boleskei, "VLSI implementation of MIMO detection using the sphere decoding algorithm," *IEEE JSSC*, vol. 40, no. 7, pp. 1566–1577, Jul. 2005.
- [3] Z. Guo and P. Nilsson, "Algorithm and implementation of the K-Best sphere decoding for MIMO detection," *IEEE JSAC*, vol. 24, no. 3, pp. 491–503, Mar. 2006.
- [4] P. Radosavljevic and J. R. Cavallaro, "Soft sphere detection with bounded search for high-throughput MIMO receivers," *40th Asilomar Conf on Signals, Systems and Computers*, pp. 1175–1179, Nov 2006.
- [5] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Math. Comput.*, vol. 44, no. 170, pp. 463–471, Apr. 1985.
- [6] B. Farhang-Boroujeny, H. Zhu and Z. Shi, "Markov chain Monte Carlo algorithms for CDMA and MIMO communication systems," *IEEE Transactions on Signal Processing*, vol. 54, pp. 1896–1909, May 2006.
- [7] B. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. on Comm.*, vol. 51, pp. 389–399, Mar. 2003.
- [8] M. Myllyla, M. Juntti and J. R. Cavallaro, "Implementation Aspects of List Sphere Detector Algorithms," to appear *IEEE Globecom*, Nov. 2007.